

09/669,680

REMARKS

Claims 1-29 are all the claims pending in the application. Claim 30 is canceled as being redundant of claim 8. Claims 1-29 stand rejected on prior art grounds. Applicants respectfully traverse these objections/rejections based on the following discussion.

I. The Prior Art Rejections

Claims 1-29 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Lantrip et al., hereinafter "Lantrip" (U.S. Patent No. 6,298,174) in view of Ruocco et al., hereinafter "Ruocco" (U.S. Patent No. 5,864,855). Applicants respectfully traverse these rejections based on the following discussion.

A. The Rejection Based on Lantrip in view of Ruocco

Applicants respectfully traverse this rejection because the applied prior art references do not teach or suggest "clustering second documents in a second dataset using said centroid seeds" as defined by independent claims 1, 20, and 23; that "said cluster generator clusters second documents in said second dataset using said centroid seeds" as defined by independent claim 8 or "clustering said second documents using said centroid seeds" as defined by independent claim 15.

The claimed invention solves the problem of finding new categories in a second data set that did not exist in the first data set, while at the same time maintaining as nearly as possible categories from the first data set as categories in the second data set. With the claimed invention, there is no requirement, and in fact it is not assumed, that the first and second data sets have any of the same data elements in them. They are allowed to have some of the same elements, for example if the second data set included of the first data set plus some additional data elements,

09/669,680

but this is in no way a requirement for the claimed invention. The claimed invention is designed in such a way as to find the similarities between the two data sets, where they exist, while at the same time finding the key differences (emerging concepts) in the second data set.

In the rejection, the Office Action admits that Lantrip fails to disclose clustering documents from a second data set. The Office Action then goes on and states that Ruocco discloses "processing in parallel second datasets based on cluster information from previous cluster vectors in order to gain the benefit of information from previous clusters to improve analysis of subsequent datasets." The section referenced in Ruocco reads as follows: "selecting a second electronic document and comparing the vector of the second electronic document with the first cluster vector to determine if the second document vector has similar characteristics...." There are two major problems with this statement in the Office Action.

The first problem is that, contrary to what the Office Action states in the objection, there is no second data set mentioned in Ruocco. The "second electronic document" referred to in the quoted section refers to the second document of the original and there is only one data set referred to in Ruocco. This becomes even more clear as you read further in Ruocco and see that each subsequent vector is treated in exactly the same way as the "second electronic document" namely it either becomes a new cluster vector or it gets assigned to an existing cluster. Thus, there is no "second data set" analogous to the second data set in the claimed invention. There is in fact only one data set in Ruocco which is processed in a single pass through to produce clusters incrementally.

The second problem is that, even assuming for sake of argument that Ruocco did disclose a second data set, processing such a dataset "in parallel" would by its very nature make it fundamentally different from the claimed invention. To process the second data set in parallel with the first would mean that neither set takes precedence. To the contrary, with the claimed invention, the first data set be completely categorized before the second data set is processed. This must be, in order to completely calculate the centroids of the categories in the first data set, which will be utilized as seeds in the categorization of the second data set. Thus, the method of parallel processing teaches away from the claimed invention.

09/669,680

Therefore, Applicants submit that the applied prior art references do not teach or suggest "clustering second documents in a second dataset using said centroid seeds" as defined by independent claims 1, 20, and 23; that "said cluster generator clusters second documents in said second dataset using said centroid seeds" as defined by independent claim 8 and "clustering said second documents using said centroid seeds" as defined by independent claim 15. Therefore, independent claims 1, 8, 15, 20, and 23 are patentable over the applied prior art references. Further, dependent claims 2-7, 9-14, 16-19, 21, 22, and 24-29 are similarly patentable, not only by virtue of their dependency from a patentable claim, but also by virtue of the additional features of the invention they define. In view of the foregoing, the Examiner is respectfully requested to reconsider and withdraw this rejection.

II. Formal Matters and Conclusion

Minor typographical errors have been corrected in the claims. For example "and" has been corrected to "in" in claims 1 and 23. In addition claim 30 has been canceled as being redundant of claim 8. These claim amendments correct minor typographical and/or word processing errors and do not relate to the merits of the claimed invention. Therefore, these claims are not intended to change the scope of the claimed invention in any way.

In view of the foregoing, Applicants submit that claims 1-29, all the claims presently pending in the application, are patentably distinct from the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

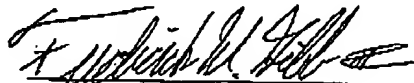
Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary.

09/669,680

Please charge any deficiencies and credit any overpayments to Attorney's Deposit
Account Number 09-0441.

Respectfully submitted,

Dated: 06/23/04


Frederick W. Gibb, III
Reg. No. 37,629

McGinn & Gibb, P.C.
2568-A Riva Road
Suite 304
Annapolis, MD 21401
Customer Number: 28211